Perspectives in Science (2016) 8, 712—714

ScienceDirect

journal homepage: www.elsevier.com/pisc

Mining of high utility itemsets of size-2 with pruning strategies and negative item values for B2C companies based on experiential marketing approach☆

Gaurav Gahlot, Nagamma Patil*

Department of Information Technology, National Institute of Technology Karnataka, Surathkal, India

Received 20 February 2016; accepted 9 June 2016
Available online 4 July 2016

KEYWORDS
High utility itemset mining;
Negative values;
Pruning;
Retail;
Offers;
Apriori algorithm

Summary Business to customer (B2C) marketing for the retail organisations is the most rapid adoption made by the developed countries, while it has the pitfalls in the developing nations. B2C companies have been continually putting some or the other lucrative offers and schemes on their manufactured products. There happens to be no logical demand of clubbing the sale of two products. The only need of such clubbing is the financial crisis which the company wants to overcome. Information technology can renew and make the competitive advantage for B2C companies. In this paper, a novel way for finding the itemset clubs is proposed, hence extending Apriori algorithm. The proposed methodology aims at finding the combinations of the products which can be sold together with the high levels of utility. This allows for making good profits for the company. Unlike contemporary way of items bearing positive values, negative item values have been looked into. The MHUIS-2wPS algorithm utilises the transactional experiences of the retail stores and outputs the size-2 clubs. The MHUI-NIV algorithm caters for the items with negative item values. The dissertation applies various pruning strategies for the discovery of high utility itemsets. These prunings will help remove the unnecessary formation of the low utility extensions. Later, various datasets have been used to show the essence of the algorithms.

Introduction

A business organisation makes a strong stand upon the decision making capabilities and its marketing strategies, along with business intelligence, warehousing, mining, querying, etc. Customer's behaviour of the retail stores is central

☆ This article belongs to the special issue on Engineering and Material Sciences.

* Corresponding author. Tel.: +91 9460993089.
E-mail addresses: gauravgahlot299@gmail.com (G. Gahlot), nagammapatil@nitk.ac.in (N. Patil).

http://dx.doi.org/10.1016/j.pisc.2016.06.067

absorbent for delivering the services, because of growing competition amongst the companies making customers fall in the dilemma to choose from. Their psychology imparts ways for making the companies lure the customers and put schemes, sales, or offers on the product items by the brands and companies. The customer's behaviour and the transaction logs have an overlapping characteristics (Homsud and Chaveesuk, 2014). All the previous researches done have support-confidence framework which uses frequency. But, the frequency may not prove to be a sufficient significant parameter since frequency is reflection of the number of transactions in the dataset containing a particular itemset, and does not imply its *Utilit*. The utility can be cost, profit, revenue or other preferred expression. Also, it might happen that frequent itemsets contribute only to a meagre part of the entire benefit, while infrequent itemsets contribute to the bigger pie of the profit. Some speculative steps have been taken in Sahoo et al. (2015). An itemset in retail business may bear negative item values. Hence, prominence arises for discovery of high utility itemsets with negative item values (Chu et al., 2009). Many supermarkets promote items to lure customers. The customers buy particular items and then get free goods which make up negative value, and ultimately supermarkets earn higher profits (Lee et al., 2013). In this paper, we propose the modifications to Apriori Algorithm called MHUIS-2wPS algorithm and MHUI-NIV algorithm. They work upon the Utility parameter to find the clubs of two item products which can be sold on schemes, using the pruning strategies for improving the performance.

High utility itemsets mining

An itemset X is a *High Utility Itemset*, if the utility of it meets the minimum threshold. Otherwise, it is a *Low Utility Itemset*. It is not efficient to apply directly the Apriori property for pruning the search space for candidate itemsets because of neither non-monotonicity nor monotonicity of the utility constraint. For reduction of the search space and enhancement of the performance, the *Transaction-Weighted Utility* (TWU) concept is used satisfying the downward closure property. TWU of an itemset X is given by $\sum_{T_q \in D, X \in T_q} tu(T_q)$.

A vertical mining approach which generates high utility itemsets in a single phase of the algorithm by the use of database's vertical representation is used. It deploys the data structure named as *Utility List* for an itemset. For an itemset X of a transaction T_q, the set of items after X in T_q is represented by T_q/X. For an itemset X, the utility list of X contains the tuples (T_q, iutil, rutil) for each transaction T_q having X. The utility of X in T_q, coined by iutil, is the sum of the utilities of all items in T_q/X. rutil is calculated as $\sum_{i \in T_q/X} u(i, T_q)$. For a given itemset X with its utility list, if the sum of iutil is less than minimum utility, then X is not a high-utility itemset.

Pruning strategies

Estimated utility co-occurrence pruning (EUCP) strategy is used to eliminate a low utility extension and also, all its transitive extensions. It encompasses the TWU pruning strategy to prune a bigger itemset of length more than two by incorporating the TWU values of all its subsets of length one. The utility value given to the estimated utility co-occurrence values by i has no effect after the subproblem of i is completed. This idea gives rise to provide a promising utility co-occurrence structure (PUCS) for every promising item for further reduction of the candidate itemsets.

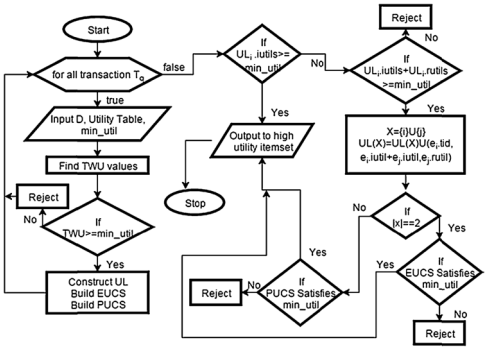

Figure 1 MHUIS-2wPS algorithm.

Proposed algorithms

The proposed MHUIS-2wPS algorithm (Fig. 1) follows the sequential approach for finding the high utility itemsets. Using the utility list, the high utility itemsets will be found. Then applying the pruning concepts of EUCS and PUCS, the itemsets will be made minimal resulting in the formation of high utility itemsets. It builds the necessary data structures and parameters for carrying out the processing. It also initiates the finding of the clubs of items. Later, it checks the other extra areas i.e. the itemset clubs which can be searched here itself for calling as high utility or not. Lastly the validation of the formed clubs is done using the decisions of EUCS and PUCS (Fig. 2).

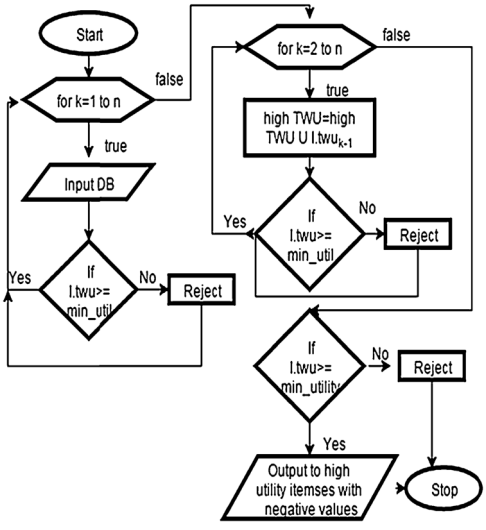

Figure 2 MHUI-NIV algorithm.

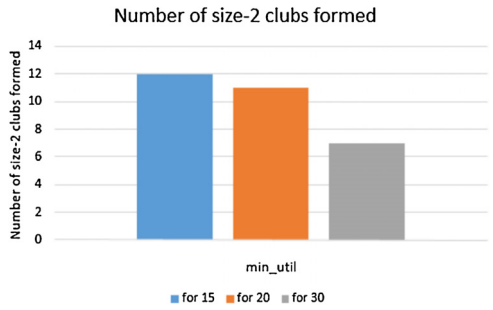

Figure 3 Number of size-2 clubs formed.

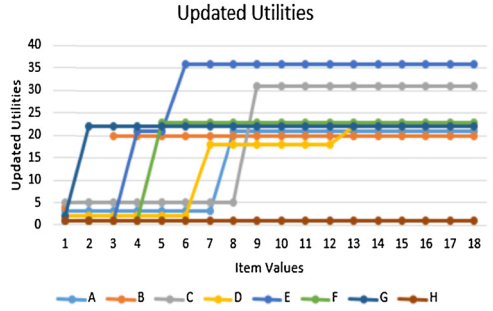

Figure 4 Updated utilities.

Results and discussions

On applying the algorithm on the synthetic dataset bearing 9 transactions, following graphs have been plotted. Fig. 3 describes the number of size-2 itemsets formed having high utility. As the iterations proceed, the item combinations are processed. The itemsets formed are checked for their utility value. Different minimum utility thresholds were taken and the graph has been plotted against the number of size-2 itemsets found. Fig. 4 depicts the way the items attain their utility gradually while executing the program. As and when the algorithm flows, the utility of the items keeps on changing. This utility as discussed is the importance value of each item. It may happen that the utility for an item may take many steps as is in the case of the item 'D' while may remain stagnant like for 'H'. This happens because of the buying nature of the customers.

We have taken in account real data, BMS-POS which contains sales data of many years of electronics retailer. Each item represents a category, instead of an individual product. Fig. 5 shows the performance upgradation for negative

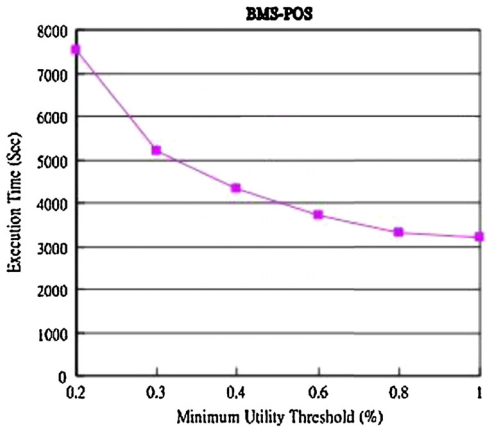

Figure 5 Results with BMS-POS.

itemsets dataset in terms of the execution times. It is noted that with the increment of the minimum threshold, the execution time decreases since the number of pruned items gets increased.

Conclusion

To provide a pragmatic relationship between itemsets, we have enforced the utility framework to mine itemsets having high utility. The algorithm, on comparison to Apriori, works faster and also give semantic relationship within the items on the basis of utility, even for negative itemsets. It is found more sustainable for making a dataset with negative values provide a better performance metrics.

References

Chu, C.-J., Tseng, V.S., Liang, T., 2009. An efficient algorithm for mining high utility itemsets with negative item values in large databases. Appl. Math. Comput. 215 (2), 767—778.

Homsud, S., Chaveesuk, S., 2014. An integrated model of customer repurchase intention in B2C e-commerce. In: 6th International Conference on IEEE Information Technology and Electrical Engineering (ICITEE).

Lee, D., Park, S.-H., Moon, S., 2013. Utility-based association rule mining: a marketing solution for cross-selling. Expert Syst. Appl. 40 (7), 2715—2725.

Sahoo, J., Das, A.K., Goswami, A., 2015. An efficient approach for mining association rules from high utility itemsets. Expert Syst. Appl. 42 (13), 5754—5778.